

Uncertainty Analysis Methods for Comparing Predictive Models and Biomarkers: A Case Study of Dietary Methyl Mercury Exposure

Rafael A. Ponce,* Scott M. Bartell,* Terrance J. Kavanagh,† James S. Woods,† William C. Griffith,* Robert C. Lee,* Timothy K. Takaro,‡ and Elaine M. Faustman†§

*Department of Environmental Health and Consortium for Risk Evaluation with Stakeholder Participation; §Center on Human Development and Disability, University of Washington Institute for Risk Analysis and Risk Communication; †Center for Ecogenetics and Environmental Health; and ‡Department of Medicine, University of Washington, Seattle, Washington 98105

Received February 6, 1998

Biologically based markers (biomarkers) are currently used to provide information on exposure, health effects, and individual susceptibility to chemical and radiological wastes. However, the development and validation of biomarkers are expensive and time consuming. To determine whether biomarker development and use offer potential improvements to risk models based on predictive relationships or assumed values, we explore the use of uncertainty analysis applied to exposure models for dietary methyl mercury intake. We compare exposure estimates based on self-reported fish intake and measured fish mercury concentrations with biomarker-based exposure estimates (i.e., hair or blood mercury concentrations) using a published data set covering 1 month of exposure. Such a comparison of exposure model predictions allowed estimation of bias and random error associated with each exposure model. From these analyses, both bias and random error were found to be important components of uncertainty regarding biomarker-based exposure estimates, while the diary-based exposure estimate was susceptible to bias. Application of the proposed methods to a simple case study demonstrates their utility in estimating the contribution of population variability and measurement error in specific applications of biomarkers to environmental exposure and risk assessment. Such analyses can guide risk analysts and managers in the appropriate validation, use, and interpretation of exposure biomarker information. © 1998 Academic Press

Key Words: biomarkers; risk assessment; methyl mercury; exposure; uncertainty.

INTRODUCTION

Biologic measures (biomarkers) are developed to provide information regarding individual exposure, effect, and/or susceptibility to a variety of environmental contaminants (Clarkson, 1987; Hattis, 1991; Albertini *et al.*, 1993; Cullen and Redlich, 1995; Grandjean, 1995;

Links *et al.*, 1995; Lowry, 1995; Rothman *et al.*, 1995; Srivastava and Rossi, 1996). These biomarkers have a range of practical applications including evaluation of toxicity mechanisms and end points, clinical medicine, and industrial hygiene (Schulte and Perera, 1993). When used in risk assessment, information from biomarkers may replace default assumptions that are made when specific information regarding exposure, absorption, toxicokinetics, and toxicodynamics is unavailable (Table 1). However, quantitative evaluations of the utility of biomarker information in risk assessments are lacking. Such assessments may be valuable for researchers and risk managers in determining whether the costs of developing and validating a biomarker are justified by the added information they would provide.

To estimate dietary exposure to an environmental agent, there are two common options: use of a predictive model and use of a biomarker of exposure. Predictive modeling usually entails examination of the major potential routes of exposure, estimation of the duration of exposure to members of a population, estimation of the concentration of the agent in each medium that contributes to exposure, estimation of absorbed fraction, and other factors. Biomarkers offer the potential for improving exposure information by allowing more specific information on individual exposure or uptake to be obtained. However, exposure estimates obtained from biomarkers applied in field studies (a noncontrolled environment) are only as good as their accuracy and precision. In a laboratory experiment or clinical trial, where a known exposure can be calibrated against the biomarker of exposure, the uncertainty around the predictive model-based exposure estimate and the biomarker-based exposure estimate can be small and the estimates highly correlated, but this may not be the case in the field where the two estimates may deviate considerably. The central question then becomes one of determining which of the exposure estimates used in uncontrolled situations (e.g., epidemi-

TABLE 1
Use of Biomarkers to Refine Risk Assessment Information

Variable	Use of biomarkers
Exposure	Establish exposure characteristics <ul style="list-style-type: none"> •Route of exposure •Peak exposure •Total exposure
Absorption	Estimate cumulative exposure Establish absorption factors <ul style="list-style-type: none"> •Inhalation •Dermal exposure •Ingestion
Toxicokinetics	Identify factors that influence absorption Identify interspecies differences Identify sensitive population characteristics Establish distribution kinetics Establish half-life in blood or body Identify interspecies differences Identify factors that influence distribution, metabolism, or excretion Estimate cumulative exposure Estimate peak exposure variables <ul style="list-style-type: none"> •Time •Concentration
Toxicodynamics	Identify sensitive population characteristics Identify mechanism of toxicity at target organ Establish target organ potency Identify sensitive population characteristics Identify factors that influence target organ toxicity Identify interspecies differences

ologic studies) deviate least from what you would expect based on data obtained through controlled studies. Using the proposed methods, we demonstrate the utility of using combined uncertainty estimates, similar to those used in regulatory risk assessment, for estimating and comparing the uncertainty provided by biomarkers of exposure and standard predictive models for dietary mercury estimation. While such analyses can be generally applied to evaluate a broad range of biomarkers of exposure, published data providing simultaneous information on dietary mercury exposure using biomarkers of exposure, self-reported dietary intake, and measured dietary intake, make the application to mercury an excellent case application.

Mercury is a potentially toxic contaminant prevalent at many hazardous waste sites including DOE and National Priority List (NPL) sites. Mercury is ranked third behind lead and arsenic on the ATSDR/EPA priority list of hazardous substances (ATSDR, 1997) based on toxicity and prevalence at contaminated sites. In the environment, elemental mercury may undergo oxidation to inorganic mercury (Hg) and biotransformation to methyl mercury (MeHg). MeHg is a naturally occurring environmental compound that bioaccumulates in fish and piscivorous species; MeHg comprises greater than 95% of total mercury in edible

fish tissues (Bloom, 1992). High exposure to MeHg during fetal development has been linked to adverse human development including mental retardation and altered motor function (Matsumoto *et al.*, 1965; Choi *et al.*, 1978). Because of its relevance to public health, biomarkers of MeHg exposure have been developed. These biomarkers include the chemical analyses of blood and hair for Hg, which have been shown to be specific and sensitive measures of exposure following prolonged MeHg exposure (Kershaw *et al.*, 1980; Phelps *et al.*, 1980; Marsh, 1987; Lind *et al.*, 1988).

Evaluating Uncertainty in Exposure Modeling

Uncertainty in exposure model variables comes from many sources including instrument error, recall bias, and population variability; these can be grouped as systematic error (i.e., bias) and random error. The effects of these errors on exposure estimates can be expressed as

$$x_i = t_i + e_i \quad (1)$$

where x_i represents the measured or estimated exposure value for the i_{th} subject, t_i represents the true exposure value for that subject, and e_i represents the measurement error for that subject (Armstrong *et al.*, 1992). Measurement bias is defined as the mean of the measurement error distribution (μ_E). The random error is described by the variance of the measurement error (σ_E^2). "Precision" is used to describe the magnitude of random error, but is inversely related, so that a highly precise measurement is one with low error variance. "Accuracy" refers to both precision and bias, with high accuracy implying high precision and low bias.

Where x_i and t_i are known for all i subjects, bias and random error can easily be assessed. However, t_i is seldom (if ever) actually known for any individual because only x_i is observed. Importantly, the individual values x_i , t_i , and e_i can be considered as observations of random variables X , T , and E , respectively. Although T and E are assumed to be independent, X is dependent on both T and E . These distinctions can be important if Monte Carlo simulation is used to simulate missing values. For example, exposure modeling using two-dimensional Monte Carlo simulation requires the estimation of the distribution of population variability T (Cohen *et al.*, 1996; Frey and Rhodes, 1996), whereas typically only a set of measured values x_i and a rough estimate of the distribution of measurement error (E) are normally available.

True differences in characteristics among members make up population variability T , and the variance of T (σ_T^2) cannot be reduced by further or improved measurement (Morgan and Henrion, 1990). Measurement error, E , represents the residual error that cannot be explained by population variability. Unlike σ_T^2 , the

variance of measurement error (σ_E^2) can be reduced by refinements in measurement and improved characterization of the variable of interest (e.g., improved characterization of the distributional form or improved characterization of the tails of the distribution). Bias exists when the mean of the measured exposures is different than the mean of the true exposures and is the difference between the true population mean (μ_T) and the mean of the measured values (μ_X ; Armstrong *et al.*, 1992). When inferences are made regarding the exposures of the broader population from which the study individuals were sampled, measurement error only accounts for one component of the potentially reducible uncertainty; we use reducible uncertainty when referring to this form of uncertainty. In this paper, we examine the effects of measurement error on estimates of T , but do not attempt to quantify the total reducible uncertainty associated with using T in inferences about broader population exposure.

In an approach proposed by Frey and Rhodes (1996), the mean and variance of the true population variability T are estimated as

$$\mu_T = \mu_X - \mu_E \quad (2)$$

$$\sigma_T^2 = \sigma_X^2 - \sigma_E^2 \quad (3)$$

Note that if the measurements are unbiased ($\mu_E = 0$), then $\mu_T = \mu_X$. While Eqs. (2) and (3) can be used to estimate the mean and variance of T , the distributional form of T depends on the specific characteristics of the distributions of X and E . One known outcome is that the distribution of T is approximately normal if the distributions of X and E are approximately normal. Similarly, if the error structure is multiplicative instead of additive, the distribution of T is approximately lognormal when the distributions of X and E are approximately lognormal. Where the normal assumption does not hold, Monte Carlo simulation of the distribution of true population variability T can be conducted. This requires the simulation of individual values of e_i conditional on x_i . Available techniques for simulating such random draws generally require the correlation of the population variability X with the measurement error E .

To demonstrate the use of uncertainty analysis applied to exposure modeling, we compare exposure estimates derived from self-reported, diary-based fish intake estimates and fish mercury content (i.e., the predictive exposure model) against exposure estimates from hair or blood mercury concentration (i.e., the biomarker-based exposure models), and we examine the factors that contribute to uncertainty about these estimates of dietary mercury intake. In these analyses, we use mercury intake estimates based on simultaneously obtained duplicate diet measurements as our reference

for assessing bias and error associated with biomarker-based or predictive model-based exposure estimates.

METHODS

To compare exposure estimates using a predictive model of exposure with blood and hair mercury analyses as biomarkers of exposure, we evaluate four exposure models, two of which are predictive models and two of which are based on biomarkers of exposure. Standard exposure equations (U.S. EPA, 1989) using fish mercury content estimates and either duplicate diet fish intake measurements (Exposure₁) or self-reported fish consumption rates (Exposure₂) are used as predictive exposure models. The biomarker-based exposure models are based on either hair Hg analysis (Exposure₃) or blood Hg analysis (Exposure₄). The four exposure models are

$$\text{Exposure}_1 \text{ (mg/day)} = [\text{MeHg}]_{\text{Fish}} \cdot \text{Fish}_{\text{duplicate}} \quad (4)$$

$$\text{Exposure}_2 \text{ (mg/day)} = [\text{MeHg}]_{\text{Fish}} \cdot \text{Fish}_{\text{diary}} \quad (5)$$

$$\text{Exposure}_3 \text{ (mg/day)} = [\text{Hg}]_{\text{Hair}} \cdot F_{\text{Blood:Hair}} \cdot F_{\text{Intake:Blood}} \quad (6)$$

and

$$\text{Exposure}_4 \text{ (mg/day)} = [\text{Hg}]_{\text{Blood}} \cdot F_{\text{Intake:Blood}} \quad (7)$$

where the factor $F_{\text{Intake:Blood}}$ is the steady-state ratio of dietary MeHg intake to Hg level in blood, and $F_{\text{Blood:Hair}}$ is the steady-state ratio of Hg levels in blood and hair. We refer to these two steady-state factors as “ F ratios.” Non-fish-related MeHg exposures are assumed to be negligible in the derivation of Exposure₃ and Exposure₄. To calculate absorbed dose, each exposure equation above would be multiplied by a gastrointestinal absorption fraction.

To evaluate the differences in the MeHg exposure estimates obtained by using biomarkers of exposure relative to predictive models, data on intake, hair Hg, and blood Hg in the same population were compiled. We used reported mercury levels in hair, blood, and diet from a study of 98 individuals with long-term, stable consumption of fish (i.e., assumed steady-state conditions; Sherlock *et al.*, 1982).

Lognormal probability distributions were fit to individual exposure data taken from Sherlock *et al.* (1982) for Fish_{diary}, Fish_{duplicate}, [Hg]_{Hair}, and [Hg]_{Blood}. While Sherlock *et al.* (1982) did not report individual [MeHg]_{Fish} values, the [MeHg]_{Fish} distribution parameters were estimated by dividing the distribution of Exposure₁, which was fit to data provided as a graph in Sherlock *et al.* (1982), by the distribution of measured fish intake (Fish_{duplicate}), which was also fit to graphed data.

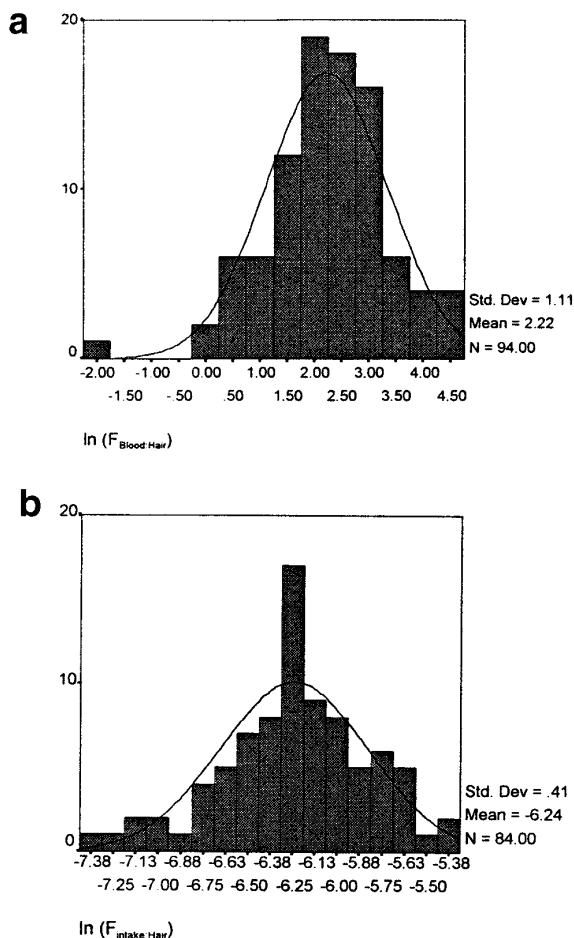


FIG. 1. (a) Lognormal distribution fit to $F_{\text{Blood:Hair}}$ ratio in Sherlock *et al.* (1982) data. (b) Lognormal distribution fit to $F_{\text{Intake:Hair}}$ ratio in Sherlock *et al.* (1982) data.

Distribution parameters for $F_{\text{Blood:Hair}}$ and $F_{\text{Intake:Blood}}$ were derived using data presented as Tables 6 and 7 in WHO (1990) and data presented in Sherlock *et al.* (1982). $F_{\text{Intake:Blood}}$ and $F_{\text{Blood:Hair}}$ were assumed to be lognormally distributed (Figs. 1a and 1b) and reported regression slopes (WHO, 1990) were used to estimate the means and variances, using weights based on sample sizes. Because the reported regression slopes did not generally include variances, the variance of each F ratio in the Sherlock *et al.* data was used as a common estimate of variance among all studies. The intent of this procedure was to develop broad-based, general estimates of the observed mean and variance of each F ratio to reflect the full range of observed values and to allow exposure modeling in study populations without prior knowledge of population-specific F -ratio values. Resulting distribution parameters for these and all other model variables are shown in Table 2.

Uncertainty propagation was conducted analytically with the assumption that all inputs within each exposure model are independent lognormally distributed

variables. Use of the lognormal distribution is appropriate for variables confined to positive values and is widely familiar to scientists and regulators (Finkel and Evans, 1987). The selection of lognormal probability distributions for model variables allows analytical uncertainty propagation after multiplicative equations are log-transformed (Morgan and Henrion, 1990). Log-normal transformation of each multiplicative equation results in a series of additive terms, such as

$$\ln \text{Exposure}_3 = \ln [\text{Hg}]_{\text{Blood}} + \ln F_{\text{Intake:Blood}} \quad (8)$$

The mean and variance for the left side of the equation are equal to the sum of the means and sum of the variances, respectively, of the variables in the right side (Fisher and Van Belle, 1993).

Using the conceptualization described by Eq. (1), alternative exposure measurements for any one subject can be viewed as various estimates of x_j . Differences in exposure predictions among the biomarker-based exposure distributions (i.e., Exposure_3 and Exposure_4) and the dietary model-based exposure distributions (Exposure_1 and Exposure_2) can be characterized by differences in the measured means to indicate relative biases ($\Delta\mu_E$) between exposure estimates and by differences in the measured variances to indicate the relative magnitude of their random errors ($\Delta\sigma_E^2$). Because σ_T^2 is defined by the true underlying frequency distribution, which is assumed to be the same for all four exposure models when applied to the same population, the difference in measured variances across models is equal to $\Delta\sigma_E^2$. If one of the models is believed to be an unbiased (but not necessarily precise) representation of T , so that $\mu_T \approx \mu_X$, then the absolute bias can be calculated for each of the other models. Similarly, if one of the models is believed to be a precise (but not necessarily unbiased) representation of T , such that $\sigma_E \approx 0$, then the absolute variance of the measurement error is known for all other models. In fact, because the model which minimizes σ_X^2 will also minimize σ_E^2 , simple inspection for the model with the lowest value of σ_X^2 will indicate the most precise model if the observed variation in each model is truly a result of independent random error affecting a common underlying MeHg exposure rate.

Relative bias and random error can be estimated between models using the results of the four exposure models [Eqs. (4)–(7)] as independent “measurements” of x_j . However, while relative bias and error provide valuable information regarding the differences between the accuracy of different exposure predictions, they do not allow determination of which of the exposure models are “better” in isolation from a standard. If the model assumptions are believed to be valid and an exposure estimate with *de minimus* bias and random error is available, then the bias and random error of

TABLE 2

Lognormal Probability Distributions for Variables Used in Propagating Uncertainty in Estimating Dietary MeHg Exposure

Variable	Units	ln mean	ln SD
Fish _{Diary} ^a	kg/day	-2.89	0.697
Fish _{Duplicate} ^a	kg/day	-3.19	0.718
[Hg] _{Hair} ^a	mg/kg	1.22	0.674
[Hg] _{Blood} ^a	mg/kg	-5.06	0.731
F _{Blood:Hair} ^b	—	-5.67	0.542
F _{Intake:Blood} ^b	kg/day	0.720	1.18
[MeHg] _{Fish} ^c	mg/kg	-1.57	0.867

^a Parameters for model input variables obtained through maximum likelihood estimation using data obtained from Sherlock *et al.* (1982).

^b Global *F* ratios derived using data from Sherlock *et al.* (1982) and WHO (1990) as described in text.

^c Fish MeHg concentration estimated from Fish_{Duplicate} and Exposure₁ reported by Sherlock *et al.* (1982).

any other exposure estimate can be estimated through comparison against this standard. As a basis for estimating absolute bias and random error, we use measured fish consumption and fish Hg content obtained from the duplicate diet [Eq. (4)] as the reference against which the other exposure estimates are compared [Eqs. (5)–(7)]. Results from these comparisons of exposure are presented in Table 3. Because a multiplicative error model is used along with lognormal random variables, the means and variances of the natural logarithm of measured exposure ($\mu_{\ln X}$ and $\sigma_{\ln X}^2$) are compared instead of the actual means and variances. Resulting estimates of the difference in log-transformed error parameters are backtransformed, yielding a ratio of means and a ratio of variances. $\Delta\mu_E$ and $\Delta\sigma_E^2$ are used to denote the differences between calculated ratios and 1 and indicate the percentage change in mean and variance due to measurement error.

RESULTS

The parameter distributions used in propagating errors in Eqs. (4)–(7) are presented in Table 2. Calculated log mean and log standard deviations for each of the exposure models are presented in Table 3 along with backtransformed values (i.e., geometric mean and geometric standard deviation).

Because the original Sherlock *et al.* data are no longer available but were graphed as individual data points in one published study, we estimated their coordinates using imaging software, and lognormal distributions were then fit to the estimated data. Histograms, P-P plots, and Q-Q plots were generated in order to visually check the fit of lognormal distributions to data for each parameter. In each case, the assumption of lognormality provided a reasonable fit to the data.

In this analysis, we assume that the duplicate diet model is the “gold standard” against which exposure estimates derived from biomarkers of mercury exposure (i.e., hair and blood Hg concentrations) and self-reported, diary-based exposure estimates are compared. Specifically, duplicate diet-based exposure estimates are assumed to be unbiased, allowing the estimation of the bias term for the remaining three exposure models. Additionally, the random error term associated with the duplicate diet model is assumed to be zero. Table 3 provides results of the four exposure models [Eqs. (4)–(7)] applied to data obtained from Sherlock *et al.* (1982). Relative bias and random error for each exposure model compared to the duplicate diet model are reported in Table 3 along with summary parameters for the resulting exposure distributions.

We find that the diary model is biased very slightly higher than the duplicate diet model, yet results in slightly lower measurement error than found with the duplicate diet model (Table 3). Of more interest are the comparisons between the biomarker-based exposure models and the dietary exposure models. Our analysis shows the hair model to be slightly positively biased when compared to the duplicate diet model ($\Delta\mu_E = 51\%$), but with a fairly large increase in random error ($\Delta\sigma_E^2 = 95\%$) when compared to the diary model. The blood model in this case is even more positively biased when used to predict dietary exposure ($\Delta\mu_E = 179\%$) and also contributes a large increase in random error ($\Delta\sigma_E^2 = 95\%$) when compared to the diary model. Each biomarker model appears to add about the same amount of imprecision to the exposure estimate, but the blood biomarker model used here contributes greater bias than the hair biomarker model.

TABLE 3

MeHg Exposure (mg/Day) Estimates: Comparison of an Exposure Model and Two Biomarkers of Exposure Using Analytical Error Propagation

	Dup diet model	Diary model	Blood model	Hair model
$\mu_{\ln X}$	-4.76	-4.45	-3.73	-4.34
$\sigma_{\ln X}^2$	1.27	1.24	1.94	1.94
Median _X	0.0086	0.0116	0.0240	0.0130
GSD _X	3.08	3.04	4.33	4.02
$\Delta\mu_E$	0% ^a	35% ^a	179% ^a	51% ^a
$\Delta\sigma_E^2$	0% ^b	-3.1% ^b	95% ^b	95% ^b

^a Bias estimates are increases in estimated mean exposure relative to the duplicate diet model, which is assumed to be unbiased.

^b Random error estimates are increases in estimated variance of exposure relative to the duplicate diet model, which is assumed to have perfect precision (i.e., no variance due to error). Note that the total calculated variance, however, is lower for the diary model than the duplicate diet model.

DISCUSSION

In this article, we apply quantitative uncertainty analysis to evaluate differences in dietary MeHg exposure estimates based on dietary self-reporting or biomarkers of exposure. This analysis provides interesting insights into the use of biomarkers of exposure and predictive modeling to estimate exposure. For example, Sherlock *et al.* (1982) reported estimates of reporting bias in the dietary recall of fish consumption. Comparison of mean reported intake (i.e., intake diaries) against mean measured dietary intake (i.e., duplicate diet) of populations from two areas demonstrated consumption overreporting by approximately 45% (0.61 kg/week vs 0.42 kg/week, respectively) and 89% (0.51 kg/week vs 0.27 kg/week, respectively). These diary-reported and duplicate diet measured intake estimates can be compared against the overall intake distributions used in the modeling described herein to determine differences in reporting bias estimates of fish intake. In our analyses, the mean diary-modeled fish intake rate is 0.39 kg/week, whereas the mean duplicate diet-modeled fish intake rate is only 0.29 kg/week (modeled diary overreporting by approximately 35%). These results suggest that the modeled fish intake reporting bias estimate (35%) is comparable, albeit smaller than, the actual reporting bias (45–89%). Comparison of actual and modeled mean fish intake rates also demonstrates that the modeled mean fish intake rates (0.29 kg/week) are at the low end of measured mean intake rates (0.27–0.42 kg/week). Because the mean is heavily influenced by the degree of variation in a lognormal distribution, these discrepancies suggest that the modeled fish intake distribution underestimates the true variance.

While individual data were available on both estimated fish intake ($Fish_{Duplicate}$) and estimated Hg intake ($Exposure_1$) in the Sherlock *et al.* study, there was no means of associating individual fish intake estimates with their corresponding Hg intake estimates. Thus, in developing the fish tissue Hg concentration distribution ($[MeHg]_{Fish}$) necessary for estimating Hg intake with the diary model ($Exposure_2$), we assumed that fish intake rates ($Fish_{Diary}$, $Fish_{Duplicate}$) and fish Hg concentration ($[MeHg]_{Fish}$) were independent, uncorrelated distributions. If such a correlation exists and it was positive, it would increase variability in both the fish Hg concentration and the exposure estimates predicted by the diary model ($Exposure_2$) relative to those reported in our analyses; a negative correlation would do the opposite.

Assuming that all error contributing to variance in the exposure measure is random and that the duplicate diet model had no measurement error (Sherlock *et al.*, 1982), we find that the diary model is the least biased relative to both the hair and blood exposure biomarkers ($\Delta\mu_E$, Table 3). However, the validity of assuming

that the duplicate diet estimate of exposure is the most accurate model of exposure depends on the measure and period of exposure being considered. For example, Sherlock *et al.* (1982) reported collection of new hair (2.4 cm) and blood samples at the end of the month-long exposure evaluation period for analysis of hair or blood Hg, respectively. Because the half-life of mercury in blood is roughly 50 days (Stern, 1997), and 2.4 cm of hair growth represents growth that occurs during slightly more than 2 months, both of these biomarkers would be influenced by exposures that occurred outside of the month-long study period; such confounding does not exist for either the diary-based or the duplicate diet-based exposure estimates. Moreover, because Hg that is incorporated into hair is in equilibrium with blood Hg at the time of hair growth, hair Hg is influenced by the half-life of mercury in blood. In this sense, both hair and blood Hg analyses are pooled estimates of exposure that are influenced by that day's exposure and, to a diminishing degree, by exposures that took place previously (WHO, 1990). Thus, our assumption (also the assumption of Sherlock *et al.*) that the periods of exposure are comparable among models and that intake is at steady state will introduce differential sampling error because a single biomarker measurement will represent a weighted average of several months of daily exposure.

With sufficient data the relative error contributed by the differences in exposure period in each biomarker-based model could be estimated by treating the biomarker as a pooled sample. Our results suggest that if the research question is focused on estimating individual exposure over 1 month, one should be cautious of using hair or blood, which will be influenced by previous exposures. However, if one is interested in determining long-term exposure trends or obtaining retrospective exposure histories, then use of hair, and perhaps blood, is likely to be a better base for estimating exposure relative to dietary histories, which may be affected to varying degrees by bias, individual recall error, and other errors (Thomas *et al.*, 1993; Bingham, 1991). When the assumption of steady-state exposure does not apply, toxicokinetic models may be used to predict tissue Hg concentrations (Suzuki *et al.*, 1993) rather than linear proportionality constants (e.g., $F_{Blood:Hair}$ and $F_{Intake:Blood}$ ratios); in such cases, the use of sophisticated modeling approaches to establish appropriate sampling protocols is likely to be required, and the most accurate medium for estimating exposure will change depending on the question being asked.

Given that both $Fish_{Duplicate}$ and $Fish_{Diary}$ are independent of $[MeHg]_{Fish}$, the differences between $Exposure_1$ and $Exposure_2$ can be attributed solely to differences in measured and reported fish intake ($Fish_{Duplicate}$ and $Fish_{Diary}$, respectively Table 2). A comparison of the variance estimated by $Exposure_2$ with $Exposure_1$ suggests that the diary-based exposure

model underestimates exposure variance relative to the duplicate diary model ($\Delta\sigma_E^2 = -3.1\%$; Table 3). This result may derive from intake underreporting among individuals with high levels of intake, intake overreporting among individuals with low levels of intake, or both. The phenomenon of systematic bias and differential over- or underreporting of intake as a function of intake level has been found to be especially troublesome when using diary-based or self-reported intake measures with populations that include certain subpopulations such as young women, athletes, or obese individuals (Bingham, 1991; Klesges *et al.*, 1995). While duplicate diet analyses have been proposed by the World Health Organization, U.S. EPA, and other agencies as the recommended method for obtaining dietary intake information, these may also suffer from both random and nonrandom errors depending on study design, length, study population, level of education, training in estimating or measuring portion sizes/weights, and other factors (Thomas *et al.*, 1997). Unfortunately, Sherlock *et al.* (1982) did not discuss the design of the duplicate diet study or factors that could influence the interpretation of the duplicate diet results.

In such cases where the assumption of uncorrelated true exposure (σ_T^2) and measurement error (σ_E^2) has been violated, one cannot assume that the exposure model with the lowest error ($\Delta\sigma_E^2$) is the most precise. Because the self-reported fish intake variance is smaller than the actual variability (as estimated by the duplicate diet; Table 3), the use of self-reported fish consumption to estimate dietary Hg exposure would underestimate population variability in exposure. The extent of this underestimation would depend on the nature of the diary-based intake reporting bias.

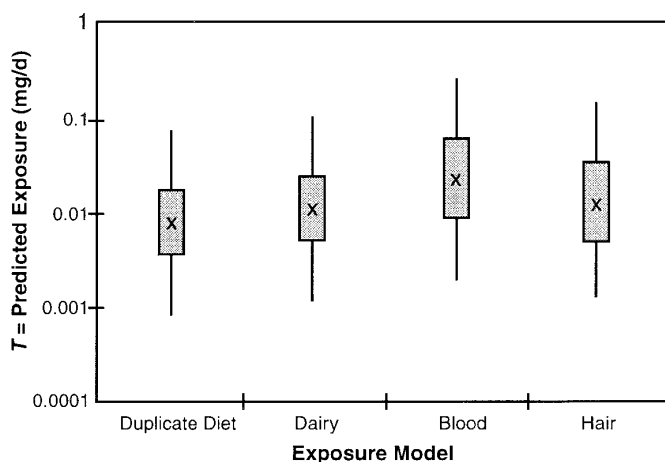


FIG. 2. Population variability in MeHg exposure estimates (mg/day) as determined by dietary exposure models (diary-based and duplicate diet) and biomarkers (hair Hg analysis and blood Hg analysis). 2.5, 25, mean, 75, and 97.5 percentile exposure estimates are shown.

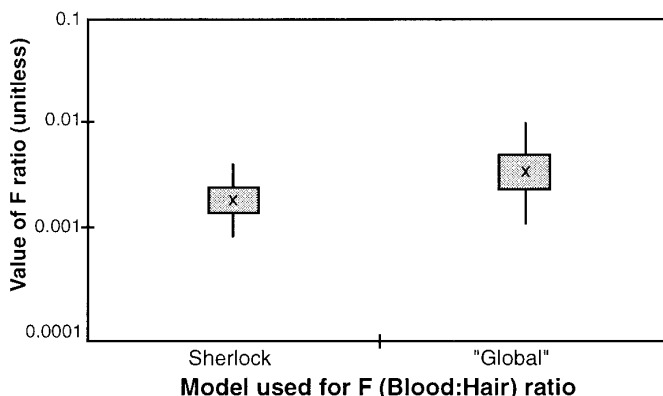


FIG. 3. Population variability of steady-state ratio of blood mercury concentration to hair mercury concentration from Sherlock and "global" estimates. 2.5, 25, mean, 75, and 97.5 percentile estimates are shown. Sherlock distribution reflects lognormal fit to observed individual values. Global distribution reflects lognormal fit to WHO reported mean values, increased by the variance observed for individuals observed in Sherlock study.

Even at this simplest level of consideration, however, little quantitative information about the relationship between the biomarker concentration (i.e., concentration of mercury in either hair or blood) and the mercury exposure rate is available in the general literature. Regulatory and scientific documents typically report the maximum likelihood estimate (MLE) of a linear regression for this variable, but fail to report regression confidence limits or other indicators of variability in the observed data; these MLE point estimates are typically used in regulatory processes to set acceptable exposure limits. To estimate the variance used to model each "global" F ratio (Figs. 2 and 3), we added the variance among MLE regression slopes reported by WHO (1990), which were weighted by sample size, and the variance among individuals within the Sherlock *et al.* study. Because individual data are not provided in the WHO document to estimate individual variability within studies, we assumed that the variability in each study cited by WHO is comparable to that of the Sherlock *et al.* study. Certainly other reasonable approaches could be taken in estimating these parameters, which would affect the outcome of this analysis. For example, Carrington *et al.* (1997) fit a series of models, weighted by goodness-of-fit, to the reported Sherlock *et al.* study individual $F_{\text{Intake:Blood}}$ values and used the resulting model variability in combination with the individual heterogeneity to estimate the overall uncertainty regarding the $F_{\text{Intake:Blood}}$ ratio. Carrington *et al.* chose not to include data from other studies in their estimate of $F_{\text{Intake:Blood}}$ due to the low correlation between dietary intake and blood mercury in those studies.

A similar approach, focusing on the Sherlock *et al.* data alone, was initially considered for developing the F ratios for this study (Figs. 2 and 3). This approach is

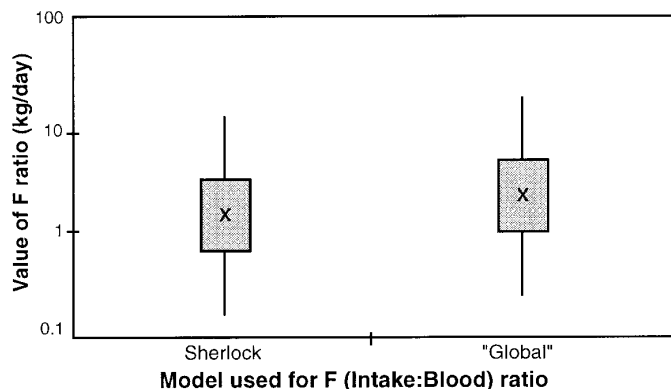


FIG. 4. Population variability of steady-state ratio (kg/day) of dietary mercury intake rate to blood mercury concentration from Sherlock and "global" estimates. 2.5, 25, mean, 75, and 97.5 percentile estimates are shown. Sherlock distribution reflects lognormal fit to observed individual values. Global distribution reflects lognormal fit to WHO reported mean values, increased by the variance observed for individuals observed in Sherlock study.

flawed in that it requires the use of the exposure data to determine the site-specific F ratios, which are then used to estimate exposure; using ratios developed in this way, we would derive the same exposure estimates for the dietary model and the biomarker model (i.e., it is not informative to use the same data to both develop and test a model). Because the purpose of using a biomarker of exposure is to replace or supplement an exposure estimate, this approach would not inform the exposure modeling process in any manner. Moreover, an exposure biomarker only has utility if its relationship to exposure estimation is in some way generalizable. Estimation of "global" F -ratio distributions (Figs. 2 and 3) is an initial attempt at generalizing this relationship and accounting for the uncertainty regarding the application of these ratios across populations. We recognize, however, that if there are regional or population differences that contribute to the reported variation in F ratios (i.e., blood:hair ratio or intake:blood ratio), then the use of local/regional F ratios would likely provide the most accurate estimates of exposure.

It is clear that the choice of F ratios strongly affects the dietary mercury exposure estimates. Moreover, in deriving and comparing global F ratios for the blood:hair partition coefficient (Fig. 3) and the intake:blood ratio (Fig. 4), we noted that there are large, potentially significant differences in the F ratios derived across populations (e.g., see WHO, 1990). Slight changes in those parameters can produce very different predictions of exposure, sometimes even altering the conclusion regarding which model produces the most precise exposure estimates.

The sensitivity of model output to slight changes in F -ratio values highlights the importance of both the assumed nature of the relationship between biomarkers and exposure when making exposure inferences

based on biomarker data and the importance of research that contributes to our understanding of the nature and reasons for variability across different populations. For example, Hg uptake into hair may vary among individuals depending on regional or genetic differences (Suzuki *et al.*, 1993). In addition, use of total hair Hg as a surrogate for MeHg intake may be confounded by exogenous inorganic Hg contamination or by incorporation of inorganic Hg into growing hair (Suzuki *et al.*, 1992, 1993); use of hair MeHg rather than total Hg has been shown to have the highest correlation with intake (Suzuki *et al.*, 1993). If regional, genetic, or physiological differences contribute to underlying variability in mercury speciation or interorgan partitioning in humans, separate F ratios may need to be developed to reflect such differences. A better understanding of the sources and magnitude of parameter variability will both inform us of and improve the predictive capabilities of biomarker-based models.

Other key assumptions that may have influenced the outcome of this analysis include the assumption of lognormality, which tends to accentuate the differences in variance across models. While the data appear to fit lognormal distributions fairly well, there is no recognized method for determining a unique probability distribution shape to represent any dataset. Additionally, assumed distributional shapes may be highly sensitive to some outlying points, thereby allowing a few unusual observations to drive the analysis. Another potential source of error was the estimation of data using software-assisted visual estimation from published graphs in the Sherlock *et al.* article. Some clusters of points were difficult to distinguish from each other and therefore may have been inaccurately estimated. While this only appeared to be a problem in a few cases out of dozens, it could have had a moderate impact on the variance estimates.

The analyses provided here highlight the need for data collection methods that retain individual information that allows evaluation of heterogeneity and correlations among variables. The data that formed the basis of this study are no longer available in their original form (Sherlock, 1997), so individual points had to be interpreted from published graphs. Data for some variables of interest such as fish Hg concentrations were not graphed and had to be estimated from the available data. These estimation procedures may have introduced additional error and lost potential correlations among individuals between their fish intake preferences (i.e., rate of consumption, fish Hg content) and Hg intake estimates; such issues could have been avoided if raw data had been available. Storage of data in public repositories could eliminate the need for this sort of estimation and allow researchers to view and interpret existing data for themselves.

CONCLUSIONS

We demonstrate the use of quantitative uncertainty analysis to evaluate the accuracy of biomarker-based exposure estimates relative to predictive models based on self-reported intake and mercury concentrations in fish. While this analysis focuses on population variability and measurement error, additional sources of uncertainty such as temporal sampling error and model error may also contribute to reducible uncertainty. Uncertainty analysis can be used to identify and quantify sources of error in risk assessment, which impacts risk-based decisionmaking regarding environmental health. Error characterization can also improve data collection and analysis decisions made in the risk assessment process, such as whether or not to use exposure biomarkers to measure exposure and how to interpret biomarker measurements. Uncertainty analysis can also be used to target critical areas for future research in order to maximize the potential uncertainty reduction of risk assessment research programs.

This approach, which relies on an assumed distributional shape (i.e., lognormal) for all model variables, allows for a straightforward estimation of uncertainty magnitude and potential sources of error in resultant exposure estimates. In cases where more information is available regarding distributional form, two-dimensional uncertainty analysis, which requires more specific information, may be employed. Two-dimensional (also called two-stage or second-order) uncertainty analysis has been proposed as a means for separately evaluating population variability and reducible uncertainty and their contribution to total uncertainty (IAEA, 1989).

If exposure biomarkers are to be successfully applied, their quantitative relationship to exposure needs to be clearly understood. For example, is the relationship between blood mercury concentrations and mercury ingestion rates well represented by a linear model or should more sophisticated toxicokinetic models be applied? Do significant population or regional differences in mercury kinetics exist or can global models be applied to the estimation of exposure from mercury biomarkers? Evaluation of these questions requires thorough quantitation of the population variability and reducible uncertainties regarding biomarker-based exposure estimation.

The analyses provided here demonstrate high demand for information required to establish well-characterized biomarkers of exposure. Of particular interest in this regard is the need for dietary intake tools that minimize bias and error and which can serve as independent measures against which to compare exposure estimates derived from predictive models and biomarkers; such a tool, the duplicate diet, was available in the study presented here. Such tools, which can

serve to provide relevant bases for comparison of exposure estimates derived from both predictive models and biomarkers of exposure, are fundamental to the validation of any exposure-estimating instrument and are the best means for obtaining estimates of relative bias and uncertainty in exposure predictions. While it is possible to examine the relative bias and uncertainty in exposure estimates and predictive models without an independent measure of exposure, such analyses are limited. Indeed, it may be argued that without performing a comparison against such an independent measure of exposure, exposure estimates based on predictive models or biomarkers remain unvalidated.

ACKNOWLEDGMENTS

This research was supported in part by the Consortium for Risk Evaluation with Stakeholder Participation (CRESP) by Department of Energy Cooperative Agreement DE-FC01-95EW55084. This support does not constitute an endorsement by DOE of the views expressed. Thanks also to Ravi Sanga for his assistance with the imaging software used for data capture. The authors also acknowledge the excellent comments and suggestions made by an anonymous reviewer.

REFERENCES

- Albertini, R. J., Nicklas, J. A., Fuscoe, J. C., Skopek, T. R., Branda, R. F., and Jp, O. N. (1993). *In vivo* mutations in human blood cells: Biomarkers for molecular epidemiology. *Environ. Health Perspect.* **99**, 135–141.
- Armstrong, B. K., White, E., and Saracci, R. (1992). Exposure measurement error and its effects. In *Principles of Exposure Measurement in Epidemiology*, pp. 49–77. Oxford Univ. Press, Oxford.
- ATSDR (1997). Top 20 Hazardous Substances: ATSDR/EPA Priority List for 1995. <http://atsdrl.atsdr.cdc.gov:8080/cxcx3.html>.
- Bingham, S. A. (1991). Limitations of the various methods for collecting dietary intake data. *Ann. Nutr. Metab.* **135**, 117–127.
- Bloom, N. S. (1992). On the chemical form of mercury in edible fish and marine invertebrate tissue. *Can. J. Fish Aquat. Sci.* **49**, 1010–1017.
- Carrington, C. D., Cramer, G. M., and Bolger, P. M. (1997). A risk assessment for methylmercury in tuna. *Water Air Soil Pollut.* **97**, 273–283.
- Choi, B. H., Lapham, L. W., Amin-Zaki, L., and Saleem, T. (1978). Abnormal neuronal migration, deranged cerebral cortical organization, and diffuse white matter astrocytosis of human fetal brain: A major effect of methylmercury poisoning *in utero*. *J. Neuropathol. Exp. Neurol.* **37**(6), 719–733.
- Clarkson, T. W. (1987). The role of biomarkers in reproductive and developmental toxicology. *Environ. Health Perspect.* **74**, 103–107.
- Cohen, J. T., Lampson, M. A., and Bowers, T. S. (1996). The use of two-stage Monte Carlo simulation techniques to characterize variability and uncertainty in risk analysis. *Hum. Ecol. Risk Assess.* **2**(4), 939–971.
- Cullen, M. R., and Redlich, C. A. (1995). Significance of individual sensitivity to chemicals: Elucidation of host susceptibility by use of biomarkers in environmental health research. *Clin. Chem.* **41**(12, Pt. 2), 1809–1813.
- Finkel, A. M., and Evans, J. S. (1987). Evaluating the benefits of uncertainty reduction in environmental health risk management. *J. Air Pollut. Control Assoc.* **37**, 1164–1171.

- Fisher, L. D., and Van Belle, G. (1993). *Biostatistics: A Methodology for the Health Sciences*. Wiley, New York.
- Frey, H. C., and Rhodes, D. S. (1996). Characterization, simulation, and analyzing variability and uncertainty: An illustration of methods using an air toxics emission example. *Hum. Ecol. Risk Assess.* **2**(4), 762–797.
- Grandjean, P. (1995). Biomarkers in epidemiology. *Clin. Chem.* **41**(12), 1800–1803.
- Hattis, D. (1991). Use of biological markers and pharmacokinetics in human health risk assessment. *Environ. Health Perspect.* **90**, 229–238.
- IAEA (1989). Evaluating the reliability of predictions made using environmental transfer models. International Atomic Energy Agency, Vienna.
- Kershaw, T. G., Clarkson, T. W., and Dhahir, P. H. (1980). The relationship between blood levels and dose of methylmercury in man. *Arch. Environ. Health* **35**(1), 28–36.
- Klesges, R. C., Eck, L. H., and Ray, J. W. (1995). Who underreports dietary intake in a dietary recall? Evidence from the Second National Health and Nutrition Examination Survey. *J. Consult Clin. Psychol.* **63**(3), 438–444.
- Lind, B., Bigras, L., Cernichiari, E., Clarkson, T. W., Friberg, L., Hellman, M., Kennedy, P., Kirkbride, J., Kjellstrom, T., *et al.* (1988). Quality control of analyses of mercury in hair. *Fres. Z. Anal. Chem.* **332**, 620–622.
- Links, J. M., Kensler, T. W., and Groopman, J. D. (1995). Biomarkers and mechanistic approaches in environmental epidemiology. *Annu. Rev. Public Health* **16**, 83–103.
- Lowry, L. K. (1995). Role of biomarkers of exposure in the assessment of health risks. *Toxicol. Lett.* **77**(1–3), 31–38.
- Marsh, D. O. (1987). Dose–response relationships in humans: Methyl mercury epidemics in Japan and Iraq. In *The Toxicity of Methyl Mercury* (C. U. Eccles and Z. Annau, Eds.), pp. 45–53. Johns Hopkins Univ. Press, Baltimore.
- Matsumoto, H., Koya, G., and Takeuchi, T. (1965). Fetal Minamata disease: A neuropathological study of two cases of intrauterine intoxication by a methyl mercury compound. *J. Neuropathol. Exp. Neurol.* **24**, 563–574.
- Morgan, M. G., and Henrion, M. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Risk and Policy Analysis*. Cambridge Univ. Press, Cambridge, MA.
- Phelps, R. W., Clarkson, T. W., Kershaw, T. G., and Wheatley, B. (1980). Interrelationships of blood and hair mercury concentrations in a North American population exposed to methylmercury. *Arch. Environ. Health* **35**(3), 161–168.
- Rothman, N., Stewart, W. F., and Schulte, P. A. (1995). Incorporating biomarkers into cancer epidemiology: A matrix of biomarker and study design categories. *Cancer Epidemiol. Biomarkers Prev.* **4**(4), 301–311.
- Schulte, P. A., and Perera, F. P. (Eds.) (1993). *Molecular Epidemiology: Principles and Practices*. Academic Press, New York.
- Sherlock, J. C., Lindsay, D. G., Hislop, J. E., Evans, W. H., and Collier, T. R. (1982). Duplication diet study on mercury intake by fish consumers in the United Kingdom. *Arch. Environ. Health* **37**(5), 271–278.
- Sherlock, J. C. (1997). Personal communication.
- Srivastava, S., and Rossi, S. C. (1996). Early detection research program at the NCI. *Int. J. Cancer* **69**(1), 35–37.
- Stern, A. H. (1997). Estimation of interindividual variability in the one-compartment pharmacokinetic model for methylmercury: Implications for the derivation of the reference dose. *Regul. Toxicol. Pharmacol.* **25**, 277–288.
- Suzuki, T., Hongo, T., Matsuo, N., Imai, H., Nakazawa, M., Abe, T., Yamamura, Y., Yoshida, M., and Aoyama, H. (1992). An acute mercuric mercury poisoning: Chemical speciation of hair mercury shows a peak of inorganic mercury value. *Hum. Exp. Toxicol.* **11**(1), 53–57.
- Suzuki, T., Hongo, T., Yoshinaga, J., Imai, H., Nakazawa, M., Matsuo, N., and Akagi, H. (1993). The hair–organ relationship in mercury concentration in contemporary Japanese. *Arch. Environ. Health* **48**(4), 221–229.
- Thomas, K. W., Sheldon, L. S., Pellizzari, E. D., Handy, R. W., Roberds, J. M., and Berry, M. R. (1997). Testing duplicate diet sample collection methods for measuring personal dietary exposures to chemical contaminants. *J. Exp. Anal. Environ. Epidemiol.* **7**(1), 17–36.
- U.S. EPA (1989). Risk Assessment Guidance for Superfund. Volume I, Human Health Evaluation Manual (Part A). Office of Emergency and Remedial Response, Washington, DC.
- WHO (1990). Environmental Health Criteria 101: Methylmercury. World Health Organization, Geneva, Switzerland.